

## University of Groningen

### Diverticular disease

Weersma, Rinse K; Parkes, Miles

*Published in:*  
Gut

*DOI:*  
[10.1136/gutjnl-2019-318231](https://doi.org/10.1136/gutjnl-2019-318231)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Final author's version (accepted by publisher, after peer review)

*Publication date:*  
2019

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Weersma, R. K., & Parkes, M. (2019). Diverticular disease: picking pockets and population biobanks. *Gut*, 68(5), 769-773. <https://doi.org/10.1136/gutjnl-2019-318231>

#### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# Diverticular disease: picking pockets and population biobanks

Rinse K Weersma<sup>1</sup> and Miles Parkes<sup>2</sup>

- 1) Department of Gastroenterology and Hepatology, University of Groningen and University Medical Center Groningen PO Box 30.001, 9700RB Groningen, the Netherlands,: [r.k.weersma@umcg.nl](mailto:r.k.weersma@umcg.nl)
- 2) Department of Gastroenterology, Cambridge University Hospital, Cambridge, CB2 0QQ, UK. [mp372@cam.ac.uk](mailto:mp372@cam.ac.uk)

**Conflict of interest** RKW received unrestricted research grants from and has acted as consultant for Takeda Pharmaceutical Company. MP has received speaker honoraria from Takeda.

**Keywords** Genome Wide Association Scan, Diverticulosis, Biobanks,

**Abbreviations:** GWAS Genome Wide Association Scan; ICD10 International Statistical Classification of Diseases and Related Health Problems.

**Word Count:** 1149

The last decade has seen an explosion in genome-wide association studies (GWAS) in many diseases. With these – and its worth reminding ourselves that this is the primary goal of GWAS - have come valuable insights into pathogenic pathways. Some have been predictable (for example variants affecting antigen presentation, co-stimulation and T-cell responses in immune mediated diseases) and others unexpected (for example autophagy in Crohn's disease). Until recently such studies had been performed in disease cohorts specifically ascertained for the purpose, with clear evidence that the larger the cohort the more loci will be found [1].

Such disease-specific cohorts with well-defined phenotypes can be assembled where there are engaged clinical communities, but can be tricky to ascertain for more general traits. Into this category falls diverticular disease. This ultimately affects up to 50% of the population but was poorly studied at a genetic level until recently, despite clear evidence for its relatively high heritability [2,3]. In the current issue of *Gut*, Schafmayer *et al* report on a large-scale analysis using clinical and genetic data from the UK Biobank. This population cohort totals 500,000 individuals among whom ~32,000 have a recorded diagnosis of diverticular disease according to ICD9/10 coding. The authors went on to replicate their findings in a hospital-based case-control cohort [4]. Interestingly, a similar approach was recently deployed by Maguire *et al*, using the *same* UK biobank 'discovery' dataset and a separate independent hospital-based registry as replication cohort [5]. The latter study identified 39 susceptibility loci for diverticular disease and showed that candidate genes reside in plausible biological pathways involved in cell-cell adhesion, membrane transport signalling and intestinal motility.

In the current paper by Schafmayer *et al* 48 susceptibility loci were identified for diverticular disease at a genome wide significant statistical threshold with consistent directionality in the discovery and replication cohorts. Of these, twelve regions were novel compared to the previous publications. GWAS loci usually harbour multiple genes and it is frequently unclear which gene is actually causal. The authors therefore performed a series of downstream *in silico* analyses to prioritize candidate genes within each locus. To the extent that these approaches have refined the associations seen, the authors highlighted one gene per locus, identified the fact that many of their signals map to introns (non-coding inserts in genes) and analysed layer-specific mRNA expression and fluorescence immunohistochemical staining in colonic biopsies. Of note, it appears that some degree of manual curation was used in the fine mapping routines, and with this the concern that some bias regarding likely causal genes might have bled into the final results. Nevertheless, novel insights into the pathophysiology of diverticular disease are derived, with the suggestion that it is a disorder of intestinal neuromuscular function, mesenteric vascular smooth muscle function and connective fibre support. These therefore overlap at least to a significant extent with the conclusions of Maguire *et al*. At this stage these mechanisms must be viewed as hypotheses to be tested. Confirmation requires more detailed genetic mapping, ascertainment of correlation between the associated genetic variants and gene expression in relevant cell types, and interrogation of their functional impact.

Now what to think of the methodology deployed in the current study? Inevitably there are some trade-offs when using a population cohort as opposed to a clinical cohort, not least in the definition of the disease under study. Schafmayer *et al* used ICD10 coding (K57) to define the case group, which includes both diverticulitis and diverticulosis. Anyone familiar with coding in the clinical setting will be aware of its potential inaccuracies, not least for a

condition such as diverticulosis where many affected individuals are asymptomatic and others may be diagnosed with the condition in the absence of objective evidence and where the true diagnosis is for example irritable bowel syndrome. Further, given the high prevalence of diverticular disease in the population, and its increase with age, one might suppose that mis-specification for 'case' and 'control' status (the latter including many 'not yet affected') would be an issue. It seems that Schafmayer *et al* used a rather more inclusive definition than the study of Maguire *et al*, and with the larger sample identified more loci at genome-wide significance. As has been well recognised previously, in genetics studies of common disease cohort size and statistical power really count, even if this comes with some loss of phenotyping accuracy.

While the potential diagnostic imprecision inherent in a population database such as the UK Biobank might be viewed as a weakness, the very fact that the cohort is broadly representative of the whole population rather than specifically collected disease cohorts itself provides an important 'real world' context for genetic findings. There is a large danger based on the published literature that clinicians view genetic results as more powerful and of greater deterministic value than they actually are. A recent publication highlights this by showing that the penetrance of causal mutations for many monogenic conditions has probably been substantially over-estimated due to their derivation from studies based on tertiary referral cohorts [6].

The UK Biobank is a prospective cohort study on ~500,000 population-based individuals [7]. For each participant a large set of phenotypes (including ICD codes as used in the current study), health related measurements, diet and lifestyle data and biomarkers are available. There are ambitious plans to link to primary healthcare records and prescribing data, as well

as possibly collecting stool samples to complement the DNA and data already collected.

GWAS has been performed on nearly the whole cohort using an array with >750,000 genetic markers. This unprecedented open access database has been made available to the research community, and – given the size of the cohort - represents a rich resource for genetic studies into common disease. For many GI disorders the numbers of affected individuals within the UK biobank is substantial. For example, >18,000 are recorded to have cholelithiasis - more than double the number of individuals included in the largest GWAS meta-analysis on gallstone disease to date [8]. Similar numbers of individuals have e.g. gastro-oesophageal reflux disease, colorectal polyps or irritable bowel syndrome. For each of these conditions there is adequate statistical power to detect genetic associations with high confidence, as nicely shown by the two studies on diverticular disease. Furthermore, other Biobanks are also coming on-stream. In the northern part of the Netherlands ‘Lifelines’, a three generations longitudinal cohort study of 165,000 participants, will have genetic data available soon, allowing for similar analyses or joint analyses with UK Biobank results [9].

These resources represent a fantastic chance for researchers to identify genetic risk factors, pathogenetic mechanisms and potential drug targets across the range of GI disease, as was done for diverticular disease. To date of the >1000 genetic studies approved by UK Biobank only a limited number relate to GI disease (see <https://www.ukbiobank.ac.uk/approved-research>). This is not good enough. Our research community needs to wake up. This is a fantastic opportunity. We should seize it!

## REFERENCES

- 1 Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat Rev Genet.* 2013;14(9):661-73.
- 2 Sigurdsson S, Alexandersson KF, Sulem P, et al. Sequence variants in ARHGAP15, COLQ and FAM155A associate with diverticular disease and diverticulitis. *Nat Commun.* 2017 Jun 6;8:15789
- 3 Granlund J, Svensson T, Olén O, et al. The genetic influence on diverticular disease--a twin study. *Aliment Pharmacol Ther.* 2012;35(9):1103-7
- 4 Schafmayer C, Harrison JW, Buch S et al. Genome-wide association analysis of diverticular disease points towards neuromuscular, connective tissue and epithelial pathomechanisms *Gut* 2019 xxxxxx
- 5 Maguire LH, Handelman SK, Du X et al. Genome-wide association analyses identify 39 new susceptibility loci for diverticular disease. *Nat Genet.* 2018 Oct;50(10):1359-1365.
- 6 Wright CF, West B, Tuke M, et al. Assessing the Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing Variants in a Population Setting. *Am J Hum Genet.* 2019;104(2):275-286.
- 7 Bycroft C, Freeman C, Petkova D et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018 Oct;562(7726):203-209.
- 8 Joshi AD, Andersson C, Buch S, et al. Four Susceptibility Loci for Gallstone Disease Identified in a Meta-analysis of Genome-Wide Association Studies. *Gastroenterology.* 2016 Aug;151(2):351-363.
- 9 Scholtens S, Smidt N, Swertz MA et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol.* 2015 Aug;44(4):1172-80